DTIC/TR-88/14

AD-A202 901

Defense Technical Information Center's Role in Numeric Database Development

Huddy B. Haller

August 1988

Defense Technical Information Center

Office of Information Systems and Technology

Cameron Station, Alexandria, VA 22304-6145

Executive Summary

Purpose of the Paper:

-demonstrate the impact of computers on numeric data

-determine whether Defense Technical Information Center (DTIC) has a role in numeric database development

-define DTIC's role

Summary:

Computers have changed the way we handle and distribute numeric data. Computers provide available information, quickly and with various data manipulations for analyses, calculations and evaluations.

DTIC has a role to play in developing numeric databases because these new resources are causing DTIC's users needs to change. DTIC 2000, the program plan, justifies DTIC's participation in numeric database efforts.

Traditional bibliographic resources that DTIC provides will no longer suffice because of the advent of computers. Numeric databases provide end users the data itself, not just pointers.

Also, DTIC users will be changing because more end users will do their own searches of numeric databases. Intermediary responsibilities will increase because they will assist end users.

DTIC's role should be to financially support and actively participate in other agencies and committees efforts. Also, DTIC's gateway system which makes multiple bibliographic databases available to its users, should make numeric databases available too. An alternative role for DTIC, developing databases, is very costly. That requires experts to collect and evaluate the data and design the database.

Recommendations:

-Fund High Temperature Materials Information Analysis Center's (HTMIAC's) online numeric database in fiscal year 1989

-Identify the type of numeric data required by DTIC users

-Participate with and sponsor organizations developing numeric databases and standards for numeric databases

-Investigate a DTIC gateway system of numeric databases

1.0 Introduction

Numeric data is numerical, textual, or graphical information.[1] The data comes from an experiment in which a scientist has measured, observed or calculated numbers, text and/or graphics.[2] Types of numeric data include physical, chemical and materials property.

Of the three types, materials property is the most important from the Defense Technical Information Center's (DTIC's) perspective. Materials provide the basis from which industry makes all products. Selecting the materials determines the cost and quality of the products. Scientists can make well-informed and reliable decisions with access to current and accurate materials data.[3]

The following statement by Dr. John Rumble of the National Bureau of Standards further describes the significance of materials property data: "Perhaps the agency with the greatest need [for materials property data] is the Department of Defense whose requirements for materials probably spans every conceivable situation."[4]

A numeric database is an organized collection of numeric data on a topic (e.g. materials property) in a computer-readable format. Magnetic tape, floppy disk for personal computer use and online system formats make database distribution possible.[5] There exists a growing need for materials property data in the online format.

This paper focuses on: the impact of computers on numeric data, whether DTIC has a role in online numeric database development, and defining that role.

2.0 The Impact of Computers on Numeric Databases

In the past, scientists have obtained the results of their colleagues' experiments from journals, periodicals and peer group contacts. Today, the personal computer has created a science and technology data explosion. Scientists and engineers increasingly use computers to handle and disseminate their most recently generated data. Scientists and engineers can electronically publish data using computers instead of incurring rapidly rising composition and printing costs.[6]

Computers have revolutionized data handling and dissemination. The value in using computers is availability of the most current information, the speed of retrieval in obtaining data and post-processing capabilities or data manipulation.[7]

Computers can store and display large amounts of data at a low cost. Telecommunications provide links between remote computers which permit the user to search multiple databases. The user can download data for further manipulation in analyses, calculations and evaluations. Graphical and statistical software packages help manipulate the data.[8]

The personal computer's popularity has increased because it makes vast amounts of data available immediately for the user to manipulate. The interest in numeric data has increased proportionally with personal computer usage. The user wants the data that numeric databases provide.

3.0 Does DTIC have a role to play in numeric databases?

DTIC must play a role in the development of numeric databases to meet its users needs. DTIC has an obligation to anticipate its users changing needs and to identify the means to satisfy their needs.

3.1 Intermediary and End Users' Changing Roles

Interest in numeric databases by DTIC's intermediary and end users is growing. Because of the advent of computers, these users' roles will be changing and DTIC must be ready to respond to their demands. Changing roles will vary depending on the institution's practices.

In institutions where end users perform the searches, their responsibilities will expand. As the number of accessible numeric databases increases, end users might often want to do their own searches for data. This most frequently occurs when they first get access or when they begin a new research and development program. If their enthusiasm fades when performing complex searches or if the novelty wears off, intermediary users can provide assistance.

Responsibilities will increase for intermediary users who understand their end users' requirements. Traditionally, intermediaries searched bibliographic databases. Now, they can help design information systems, become online consultants or members of research teams to meet other members' information needs.[9]

3.2 Numeric Databases vs Bibliographic Databases

DTIC has always provided the scientific and technical community with reliable bibliographic resources. That is, DTIC has met the needs of its users by providing pointers to the desired data's location. Numeric databases, however, provide a new dimension in information technology, the data

itself.

Bibliographic searches are more time-consuming than numeric database searches. The user searches a bibliographic database for pertinent citations. From those abstracts, the user requests reports that appear desirable. Once received and reviewed, the reports may or may not contain the data needed. If the reports do not contain the desired information, it may be necessary to conduct additional searches and to request other reports. These delays can cost scientists and researchers time and money. Often, time is critical; the request for data requires an immediate response. Online numeric database searches satisfy this need by providing the data immediately. For this reason alone, DTIC should encourage the rapid development of numeric databases.

3.3 Mandated to Provide Users with Additional Resources

DTIC 2000, a plan which provides the program direction, mandates DTIC's direct involvement in developing numeric databases. DTIC 2000 supports this mission as stated in the following long range goals:

> DTIC 2000 Goals--Future Operating Environment Goals:
> - 1. DTIC will approach the future as an innovative organization. 1.2 Actively seek new missions and information service responsibilities.
> - 3. DTIC will expand coverage of its data collections. 3.3 Devise a program for acquiring new data bases of interest to DTIC's user community and 3.4 Establish DTIC as the DoD point of entry to DoD and interagency data bases.
>
> DTIC 2000 Goals--Products and Services Goals:
> - 4. DTIC will be an information-oriented organization providing a wider range of information for DoD. 4.6 Develop interorganizational data base links.
> - 5. DTIC will be a user-oriented organization. 5.2 Enhance end-user access.[10]

4.0 Determining DTIC's Role

To perform its mission (see 3.3) and to continue its user-responsive posture, DTIC must plan a future course to provide needed data. DTIC's options include developing numeric databases or increasing its activities in investigating and encouraging numeric database efforts by other organizations.

4.1 Developing a Numeric Database:

Data is the heart of the scientific process. Data generators (original researchers) measure and report data to gain knowledge and disseminate data. At times, data generators use another person's data because of the cost and effort involved in repeating the experiment.

Also, people who are not technical experts in generating data, rely on data generators' results. These end users apply the data to experiments for calculations, analysis and design. Because they do not have a background in data validation, they look to data evaluators to verify the accuracy of the data.[11]

The Information Analysis Centers (IAC's) collect data from data generators, evaluate the information and disseminate it from a central point. Since they often act as an intermediary between the data generator and the end user, the IAC's are logical database builders. They can identify a select but active user group's well-stated demands that might provide a basis for a numeric database developmental effort. As the need for number-related information grows, these centers will play a greater role in developing and maintaining numeric databases in-house.[12]

It is very costly to develop a database and make it available in a computer-readable format. First, it is labor-intensive to identify important data from primary and secondary literature and raw data sources. The scientists manually extract numbers for evaluation purposes. Secondly, experts must evaluate the validity of the data. Accuracy limits are difficult to convey on a database. Next, the experts describe conditions under which original researchers generated data (instrumentation, calibration, use of standards, and algorithms). Scientific and technical data representation in the form of superscripts, footnotes, Greek and other alphabets, complicate database design. Then the developer designs the database by considering standardizing data for maintenance and distribution purposes. Software development costs are high for maintaining and updating data and for developing user friendly routines. The software should contain validation checks for accurate data entry.[13]

4.2 Supporting other numeric database efforts

4.2.1 HTMIAC's Online Numeric Database

DTIC has awarded HTMIAC $137,408 for fiscal year (FY) 1988. They put their high temperature materials numeric database online.[14] HTMIAC is one of the few organizations

that is doing serious numeric database developmental work.

a. HTMIAC's Successes

The High Temperature Materials Database:

During their contract award period, HTMIAC has successfully entered online almost all of its unclassified data sets on high temperature materials and properties.

They have begun the process of identifying and contacting the potential users of the database, i.e. both government employees and contractors.[15]

HTMIAC's Other Efforts:

HTMIAC not only collects data generated by reputable sources but they perform critical analysis and synthesis on raw data. Only 22% of this IAC's bibliographic holdings are DTIC's. The remainder are reports and journal articles from sources other than the government or their contractors. HTMIAC gets requests for data it collected between 1960-1985 (known then as Thermophysical and Electronic Properties IAC or TEPIAC). It covered a much wider range of materials (1000) than what HTMIAC now provides. HTMIAC is no longer collecting and evaluating this data but makes it available to DoD and their contractors. The data is still so valued, respected and complete that one-half of their requests are for this information. The staff has also taken the initiative to do other developmental work. They are developing a numeric database funded by Electric Power Research Institute.[16]

b. HTMIAC's Problems

DTIC awarded funds to HTMIAC for FY 1988 and performed an evaluation of HTMIAC's efforts in June 1988. The most significant problem has been usage of HTMIAC's database.
HTMIAC intended to make the online numeric database self-supporting after the contract award period by charging users subscription fees. Initially, the HTMIAC made their online numeric database available to DoD employees only, the Laser Hardened Materials and Structures Group (LHMSG), who were not the end users of the data. Once they got approval to make the database available to those other than DoD employees, DoD contractors, interest increased. This happened so late in the contract period that many of these users have not yet accessed the online system. Therefore, HTMIAC cannot rely on subscription fees to fund the continuing effort in FY 1989.[17]

4.2.2 Other Numeric Database Efforts in Standardization and Development of Numeric Databases

Much work in the numeric database arena is in standardizing and developing numeric databases. Standardizing aids database developers to ensure compatibility with other sources. Developing numeric databases involves collecting and evaluating data and designing the database. The following are online numeric database efforts that DTIC should track:

Government-supported efforts:

a. National Academy of Sciences, Numerical Data Advisory Board (NDAB)

NDAB's objective is to improve quality, reliability, availability, accessibility, distribution, usage and management of data. Their mission is to study or examine emerging technologies and to develop standards for numeric databases.[18]

b. National Bureau of Standards, Standard Reference Data Program

They provide reliable, well-documented data to scientists and engineers to use in decision making, research and development. They critically evaluate results from experiments and calculations. Publications and databases provide the format for their critically evaluated data. The program includes data evaluation centers in major areas of physics, chemistry and materials science. They also participate in cooperative data projects. Typically, NBS works with an outside group such as Materials Property Data Network (MPD Network) to develop databases.[19]

c. ASTM E49 (American Society for Testing and Materials Committee E49 on Computerization of Materials Property Data)

They develop standard classifications, guides, practices and terminology to build and access materials property databases. ASTM E49 is responsible for standardizing data entry, storage and interfaces.[20]

d. Versailles Project on Advanced Materials and Standards (VAMAS)

The purpose of the VAMAS Task Group on Factual Materials Databanks is to survey factual materials databanks. Based on the survey results, they identify areas where developing standards would help in building and disseminating materials databases. This group assists various agencies in developing internationally accepted standards for advanced materials.[21]

e. A Report to the Interagency Panel on Materials Data lists available and developmental materials databases. The U.S. government and other non-profit organizations fund these efforts.[22]

Commercially-Available Efforts:

The high cost of initial data collection has limited private industry's efforts in database development for profit. The lack of measurable economic benefits is industry's barrier to building a scientific and technical database. However, once government and other non-profit organizations create the initial base, industry will maintain and provide access to it.

Numeric Database Gateway Efforts:

Available materials databases lack compatibility in software, user interfaces, terminology and data representation. Often, scientists and engineers require more than one source of information. Cumbersome access makes their searches too complex and time-consuming thereby, reducing their usage.[23]

Therefore, the scientific and technical community needs a gateway system of numeric databases. It would provide users with one resource to access multiple databases. The gateway provides a means to gain easy access to shared data thereby reducing duplicative research efforts.

A gateway system includes a central computer linked via telecommunications to interconnect users (with different personal computers and terminals) to databases on mainframes in different locations. A phone call to the central computer provides a single access point to the system. The user identifies the desired data and the system provides a list of databases. The user selects the database and connects to it. The gateway provides centralized access using one command language.[24] The system has a global directory to resources, graphics, statistical packages, extraction of data and tools for downloading, analysis and uploading text and data and disseminating by electronic mail.[25]

a. Materials Property Data Network (MPDN)

MPDN's mission is to provide easy online access to numeric material property data from world wide sources. They are developing direct access to existing data services and directories and other leads to newly developing sources. This group is working to create databases, providing standards, and presenting and exchanging data.

National Bureau of Standards, Department of Energy, Lawrence

Berkeley Laboratory and Stanford University's computer technology developed the MPD Network.[26]

b. Defense Gateway Information System (DGIS)

DGIS is an information system that allows scientists, engineers, and information specialists remote access to multiple information sources like bibliographic databases. It allows the individual to log onto certain government and commercial online systems and to manipulate the data with post-processing capabilities.[27]

The intent is to provide remote access to multiple numeric databases as well but no work has yet begun in that area.

5.0 Summary

DTIC provides its intermediary and end users bibliographic databases or, pointers to the data. The popularity of personal computers has sparked interest in numeric database usage by the scientific and technical community. Computers provide access to data which the user can manipulate. DTIC's end users want what numeric databases provide: the data itself. These databases are developing slowly for two reasons:

First, development and format costs for numeric databases are very high. The developer must have the ability to collect and evaluate the data. Then, the developer must design the database, decide its format (i.e. floppy, magnetic tape, online) and enter the data.

Another reason is that numeric databases may require intermediary and end user roles to change. In some institutions, end users may do their own searches and in others, intermediary users may handle the data.

DTIC must prepare to meet the challenge of the future by defining its users' needs and the means to meet their needs. DTIC has established goals in DTIC 2000 which encourages numeric database development. DTIC's options are to design numeric databases and to support efforts of other organizations to be more responsive to its users.

Designing numeric databases requires scientific expertise which DTIC does not have. DTIC can, however, support numeric database efforts through funding and active participation in the activities of other government agencies, and other non-profit and commercial organizations. Many organizations concentrate on developing numeric databases and standards for such databases. One effort DTIC has supported is HTMIAC's

online numeric database. By doing so, DTIC supports its users because nearly 75% of HTMIAC users are DTIC users.[28] By assisting HTMIAC and other database builders, DTIC provides better and more varied services to a broader based user community.

6.0 Recommendations

DTIC must take on responsibilities for numeric database development to be responsive to its users. DTIC should actively support numeric database efforts by assuming some associated costs as shown below.

HTMIAC's High Temperature Materials Online Numeric Database:

Because HTMIAC's efforts are so near fruition, the author recommends that DTIC fund HTMIAC's efforts in the high temperature materials online numeric database for FY 89. In the project statement, Evaluating and Developing Scientific Numeric Databases, the author requested that DTIC award HTMIAC $80,000. These funds would support the online system and allow HTMIAC to make software enhancements.

In FY 89, DTIC should work with HTMIAC to do the following:

-establish timeframes to complete specific tasks

-design a fee schedule that might include reduced fees to early supporters

-attend conferences and seminars to make presentations and meet with organizations that are active in standardization and gateway efforts.

DTIC Efforts in Developing and Supporting Numeric Databases:

In the project statement, Evaluating and Developing Scientific Numeric Databases, the author requested that DTIC fund developing and supporting numeric databases for: $20,000 in FY 89, $25,000 in FY 90, and $25,000 in FY 91. DTIC would use these funds for the following activities:

-identify its users needs

-support and sponsor IAC's and other organizations whose activities include collecting, evaluating, generating, and disseminating computer-readable numeric data. Also, support and sponsor those that publish hardcopy numeric databases to automate the data.

-investigate how to provide its users with a gateway system for numeric databases.

Biblography

Chen, Ching-chih and Hernon, Peter, eds. Numeric Databases. Norwood, NJ: Ablex Publishing Corp, 1984.

DGIS Workbook: Gateway User Support and Training Office. Arlington, Va.: BBN Laboratories Inc., [May 1987].

Douglas, Richard D.; McCauley, Ellen V.; Kuhn, Allan D. et al. DTIC 2000, A Corporate Plan for the Future. Alexandria, Va.: Defense Technical Information Center, July 1984.

Haller, Huddy. "Evaluation of High Temperature Materials Information Analysis Center Online Numeric Database". Alexandria, Va.: Defense Technical Information Center, June 1988.

Hampel, Viktor E., Grubb, David P. and Moulik, Amal. "Interactive Access to Scientific and Technological Factual Databases Worldwide". Lawrence Livermore National Laboratory for U.S. Department of Energy under contract no. W-7405-ENG-48, [October 1987].

Ho, C. Y. High Temperature Materials Information Analysis Center, CINDAS/Purdue University. Numerous telephone conversations and meeting with Huddy Haller, Defense Technical Information Center, May 9-10, 1988.

Hunt, Richard K.; Fisher, H. Leonard; Hampel, Viktor E., et al. "The 'TIS' Intelligent Gateway Computer, An Alternative to the 'Doomsday Scenario'." Paper presented at 4th National ONLINE Meeting at the Sheraton Center, New York, N.Y., 12-14 April 1982. Lawrence Livermore National Laboratory for U.S. Department of Energy under contract no. W-7405-ENG-48, February, 1983.

Kaufman, J. Gilbert. "Towards Standards for Computerized Material Property Data and Intelligent Knowledge Systems". ASTM Standardization News, (March 1987).

Krockel, H. et al. eds., Factual Materials Databanks: The Need for Standards. Gaithersburg, MD: National Bureau of Standards, [July 1987].

Lide, David R. Jr. "Critical Data for Critical Needs." Science 212 (June 19, 1981):1343-1349.

Lide, David R. Jr. and Molino, Bettijoyce B. "Critical Data in Physics and Chemistry." Encyclopedia of Physical Science and Technology 3 (1986):786-791.

Numerical Data Advisory Board. Washington, D.C.: Numerical Data Advisory Board, [1987].

Purdue University, CINDAS, "A Proposal on Pilot Study of On-Line Operation of a Computerized High Temperature Materials Properties Data Base for Laser Hardening Applications." West Lafayette, Ind.: CINDAS/Purdue University, 1987.

Rumble, John ed. Advanced Materials Databases: A Report to the Interagency Panel on Materials Data. Work done for Wright-Patterson Air Force Base, Ohio Work Order # F 1457-87-N-5046 Gaithersburg, Md.: National Bureau of Standards, [1988].

Rumble, John Jr.; Mitchell, M.R.; and Northrup, R. Thomas, eds. Computerized Materials Data-A Workshop for Ground Vehicle Engineering. Warrendale, Pa.: Society of Automotive Engineers, Inc., 1985.

Rumble, John; Sauerwein, Joan; and Pennell, Sharon. Scientific and Technical Factual Databases for Energy Research and Development: Characteristics and Status for Physics, Chemistry, and Materials. Report No. DOE/TC/40017- 1 (DE87001518). Gaithersburg, MD: National Bureau of Standards, October, 1986.

Standard Reference Data Program. Gaithersburg, Md.: National Bureau of Standards.

Westbrook, J.H. and McCreight, L.R. eds. Computerized Aerospace Materials Data: Proceedings of a Workshop on Computerized Property Materials and Design Data for the Aerospace Industry. New York, NY, 1987.

Westbrook, J.H. gen. ed. Materials Data Systems For Engineering: The Proceedings of a CODATA Workshop held at Schluchsee, The Black Forest, Federal Republic of Germany September 22-27, 1985. West Germany, 1986.

Footnotes

1. David R. Lide Jr., "Critical Data for Critical Needs," Science 212 (June 1981):1343.

2. Ching-chih Chen, Numeric Databases, eds. Ching-Chih Chen and Peter Hernon (Norwood, N.J.: Ablex Publishing Corporation, 1984), p. 2.

3. John Rumble, Jr., M.R. Mitchell and R.Thomas Northrup, eds., Computerized Materials Data-A Workshop for Ground Vehicle Engineering: Report of a Workshop held April 1984 in Columbus, Ohio (Warrendale, Pa.: Society of Automotive Engineers, 1985), pp. 25-33.

4. John Rumble, ed., Advanced Materials Databases: A Report to the Interagency Panel on Materials Data (Gaithersburg, Md: National Bureau of Standards, [February 1988]), p. 1.

5. David R. Lide Jr., "Critical Data for Critical Needs," Science 212 (June 1981):1343.

6. Ibid., p. 1347.

7. J.H. Westbrook, gen. ed., Materials Data Systems For Engineering: The Proceedings of a CODATA Workshop held at Schluchsee, The Black Forest, Federal Republic of Germany September 22-27, 1985 (West Germany), pp. 89-91.

8. John Rumble, Joan Sauerwein and Sharon Pennell, Scientific and Technical Factual Databases for Energy Research and Development: Characteristics and Status for Physics, Chemistry, and Materials (Springfield, Va.: U.S. Department of Commerce, National Technical Information Service, 1986), pp. 4-6.

9. Richard K. Hunt, H. Leonard Fisher, Viktor E. Hampel, et al., "The 'TIS' Intelligent Gateway Computer, An Alternative to the 'Doomsday Scenario'", paper presented at 4th National ONLINE Meeting at the Sheraton Center, New York, N.Y., 12-14 April 1982, pp. 2-3.

10. Richard D. Douglas, Ellen V. McCauley, Allan D. Kuhn, et al., DTIC 2000, A Corporate Plan for the Future, (Alexandria, Va.: Defense Technical Information Center, July 1984), pp. 6-8 and 7-10.

11. Gesina C. Carter, Numeric Databases, eds. Ching-Chih Chen and Peter Hernon (Norwood, N.J.: Ablex Publishing Corporation, 1984), pp. 16-20.

12. John Rumble, ed., Advanced Materials Databases: A Report to the Interagency Panel on Materials Data, 11.

13. David R. Lide, Jr. and Bettijoyce B. Molino, "Critical Data in Physics and Chemistry," Encyclopedia of Physical Science and Technology 3 (1986):795.

14. C. Y. Ho and H. H. Li, "A Proposal on Pilot Study of On-Line Operation of a Computerized High Temperature Materials Properties Data Base for Laser Hardening Applications" (Request For Proposal, Purdue University/CINDAS, 1987 and basic contract DLA900-86-C-0751).

15. Huddy Haller, "The Evaluation of High Temperature Materials Information Analysis Center Online Numeric Database", (Alexandria, Va: Defense Technical Information Center, June 1988), p. 3.

16. Interview with Dr. C. Y. Ho, High Temperature Materials Information Analysis Center, W. Lafayette, Indiana, 9-10 May 1988.

17. Huddy Haller, "The Evaluation of High Temperature Materials Information Analysis Center Online Numeric Database", 11-12.

18. Numerical Data Advisory Board, (Washington, D.C.: Numerical Data Advisory Board, July 1987).

19. Standard Reference Data Program, (Gaithersburg, Md.: National Bureau of Standards).

20. J. Gilbert Kaufman, "Towards Standards for Computerized Material Property Data and Intelligent Knowledge Systems", ASTM Standardization News (March 1987):38-39.

21. H. Krockel, K. Reynard and J. Rumble, eds., Factual Materials Databanks: The Need for Standards, (Gaithersburg, Md.: National Bureau of Standards, [July 1987]), p. v..

22. John Rumble, ed., Advanced Materials Databases: A Report to the Interagency Panel on Materials Data, 16-35.

23. H. Krockel, K. Reynard and J. Rumble, eds., Factual Materials Databanks: The Need for Standards, 1.

24. John Rumble, Jr., M.R. Mitchell and R.Thomas Northrup, eds., Computerized Materials Data-A Workshop for Ground Vehicle Engineering: Report of a Workshop held April 1984 in Columbus, Ohio, 1-2.

25. Viktor E. Hampel, David P. Grubb and Amal Moulik "Interactive Access to Scientific and Technological Factual Databases Worldwide", paper presented at The Metallurgical Society, Cincinnati, Ohio, 12-14 October 1987, p. 3.

26. J.H. Westbrook and L.R. McCreight, eds., Computerized Aerospace Materials Data: Proceedings of a Workshop on Computerized Property Materials and Design Data for the Aerospace Industry, (New York, N.Y.: American Institute of Aeronautics and Astronautics, Inc., January 1987), pp. 49-61.

27. DoD Gateway Information System (DGIS) Workbook: Gateway User Support and Training Office, (Arlington, Va.: BBN Laboratories, Inc., May 1987), pp. a-2-6.

28. Huddy Haller, "The Evaluation of High Temperature Materials Information Analysis Center Online Numeric Database", 7.